DOCUMENT RESUME

ED 235 461 CS 007 237

AUTHOR Heerman, Charles E.
TITLE A Review of Traditional Cloze Testing Methodology.
PUB DATE 4 May 83
NOTE 18p.; Paper presented at the Annual Meeting of the International Reading Association (28th, Anaheim, CA, May 2-6, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Information Analyses (070)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cloze Procedure; Elementary Education; Literature Reviews; *Psycholinguistics; Readability; *Reading Research; Reading Tests; *Scientific Methodology; Test Construction; Test Reviews; *Test Validity

ABSTRACT

To analyze the validity of W. L. Taylor's cloze testing methodology, this paper first examines three areas contributing to Taylor's thinking: communications theory, the psychology of speech and communication, and the theory of dispositional mechanisms--or nonessential words--in speech. It then evaluates Taylor's research to determine how he established his initial methodology and what he later did to refine it. After reviewing other studies challenging the efficacy of the traditional cloze testing procedure, the paper concludes that, despite some reservations, the traditional methodology appears to provide a valid measure of readability. (MM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED235461

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

A Review of Traditional Cloze Testing Methodology

A Paper Presented as a Part of the Symposium:

NEW DIMENSIONS IN CLOZE TESTING: PSYCHOLINGUISTIC, MEASUREMENT, AND FUNCTIONAL PERSPECTIVES

Twenty-Eighth Annual Convention of the
International Reading Association
Anaheim, California 92803
May 4, 1983

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Charles E. Heerman
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Charles E. Heerman
Associate Professor
College of Education
Kansas State University
Manhattan, Kansas 66506

CS 007237

A Review of Traditional Cloze Testing Methodology

In reviewing traditional cloze testing methodology, it is necessary to examine the assumptions under which Taylor (1953) constructed his original testing method. Second, it is important to examine Taylor's work itself in order to see how he established the initial methodology and what he subsequently did with it. Third, some research will be touched upon which illustrates the traditional methodology to be commonly used, but open to serious question.

The Traditional Cloze Testing Methodology

When constructing a traditional cloze exercise, a passage of approximately 250 to 500 words is selected from regular classroom reading material. No deletions are made in the first or last sentence. A beginning deletion point is randomly chosen from among the first five words of the second sentence. Every-nth word thereafter is automatically and mechanically deleted and replaced with a standard length blank space until a minimum of 50 deletions are made. The student's task is to read the mutilated passage and attempt to predict the deleted word by using available context clues. When the cloze is used for testing purposes, only exact word matches of the original text are scored as correct.

Taylor's Methodological Assumptions

In 1953, Taylor formally published the findings from his first cloze research. By examining this article and other sources one is able to infer three areas of study from which Taylor drew in constructing his cloze testing methodology:

1. Communications theory (Shannon and Weaver 1949, Rankin 1978, p. 149).
2. Psychology of speech and communication (Miller 1951, Taylor 1953).
3. Dispositional mechanisms (Taylor 1953, Osgood 1952).

Communications Theory

From communications theory, Shannon (1948) and Shannon and Weaver (1949) provided three major concepts which Taylor either brought to, or, considered in his cloze testing methodology: a) the logarithmic nature of message reproduction b) entropy c) redundancy.

Figure 1 is a general communications system depicted by Shannon and Weaver (1949). With this system the message originates at the source and

---

Figure 1 goes about here

---

is transmitted (encoded), received and decoded at its destination. Shannon (1949) pointed out that the accuracy and amount of information received should be ascertained by a logarithmic measure. On this point Shannon relied on the work of Hartley (1928, 535), in noting:

> . . . the actual message is one selected from a set of possible messages. The system must be designed to operate for each possible selection, not for the one which will actually be chosen . . . If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced, when one message is chosen from the set, all choices being equally likely (p. 3).

The communication system was designed, then to deal with random sets of information in the communications channel. This randomness imposed entropy or "uncertainty" (Shannon 1949, p. 22) on the accuracy of the message at various points in the system as diagrammed in Figure 1. Shannon stated that information transmission did not have to be exact, but, that the transmission must be made to within a certain tolerance (p. 74). Returning to Figure 1, it can also be seen that noise interference could occur in the communications channel and that this interference is in addition to the entropy or uncertainty in the system.

Based on the above, Shannon posed a corrections system for controlling entropy and noise interference. Figure 2 represents Shannon's correction

---

Figure 2 goes about here

---

system which requires an observer to monitor the message (M) at the transmission point and, again, at the reception point (M'). Based on differences between M and M', the observer would forward correction data in order to recapture the essence of (M) the original message. In addition to this correction system, Shannon suggested including a certain amount of redundancy in the transmission of the message to control for the effects of entropy and noise interference. For the English language, Shannon noted that redundancy was a natural phenomena, with a statistical probability which approached randomness (p. 43).

At the same time, his co-writer Weaver (pp. 103-104) emphasized the need to maximize entropy to maximize information. He equated entropy with information noting that, ". . . information, in communication theory,

is associated with the amount of freedom of choice we have in constructing messages" (p. 103). Borrowing Shannon's measure of relative entropy (ratio of actual to maximum entropy), Weaver (p. 104) then specified the following formula for computing redundancy:

1 - relative entropy = redundancy

Weaver further suggested that the spurious measure of noise could be controlled by calculating all possible entropies which would then become an equivocation or an average entropy of the message. Later, Shannon (1951) attempted this with letters and stated:

> . . . in ordinary literary English, the long range statistical effects (up to 100 letters) reduce the entropy to something of the order of one bit per letter, with a corresponding redundancy of roughly 75%. The redundancy may be still higher when structure extending over paragraphs, chapters, etc. is included. However, as the lengths involved are increased, the parameters in question become more erratic and uncertain, and they depend more critically on the type of text involved (pp. 50 - 51).

In summary, Shannon verified, mathematically, a model for sending and receiving information in a communication system. He specified that the system be governed by logarithmic measures. He verified that redundancy and entropy were antithetical, but necessary components in this communications process. Together, Shannon and Weaver provided a fundamental direction for cloze testing methodology, by specifying the need for an appropriate mix of certainty (redundancy) and uncertainty (entropy) with adequate controls for noise interferences. Yet, it should be understood that these writers were addressing communication in the broadest

sense as ". . . procedures by which one mind may affect another" (p. 95). This included not only language, but music, the arts, theater, and telecommunications. It remained for others to refine this theory futher to focus more closely on printed language.

Psychology of Speech and Communications

In his textbook for a psychology of speech and communication course, Miller (1952) acknowledged Shannon's definition of the quantity of information being the logarithm of the number of alternatives (p. 46), and reiterated the point that by increasing the number of alternatives in a message the amount of information is increased. Using decibels as the measure, Miller illustrated when noise interference (measured in decibels) exists in the communication channel, the signal (also measured in decibels), can still be heard if the signal is proportionately loud enough (pp. 56-57).

According to Miller, if an auditory message is systematically interrupted the message can still be understood. Here he gave an example of clipping sound waves. Miller also applied this principle to context use in connected discourse:

> When larger units - phrases or sentences - are used, the listener is set to supply certain blanks in certain ways according to the words preceding and following the blanks . . . Whole words can often be replaced correctly when they are missing from a sentence. In connected discourse, we are bound by syntactic rules. Verbs, adjectives, nouns, etc., appear in certain standard patterns in English. If enough of the discourse is perceived to reveal the basic pattern of the sentence, the range of possible words that can be substituted into the pattern is greatly decreased (p. 77).

Miller also illustrated frequency distributions for letters, phonemes, consonants, syllables, and words. Through illustrations of zero- through fifth-order approximations using words, Miller further demonstrated, ". . . the extent to which any particular verbal unit is determined by the other verbal units around it" (p. 81). Therefore, it can be seen that Miller was able to more fully bridge the mathematical notions of Shannon and Weaver into the realm of printed language, and provide a more descriptive rationale for cloze testing methodology.

Dispositional Mechanisms

It was this area which Taylor (1953, 418-419) attributed to the work of Osgood (1952), and, which Taylor was most explicit about its having had an influence on cloze testing methodology. Dispositional mechanisms, according to Taylor, consisted first of "redundancy" in the English language. He reduced, "A man is coming now," to "man coming," to illustrate that the redundancy could be removed and meaning retained. He further noted, that even though language was redundant, some words are more likely to occur in sequence than others, a fact which he called, "transitional probabilities." It was, therefore, language redundancy and transitional probabilities, together, which allowed a person to develop, "dispositional language habits":

> When words come in sequences that best fit the existing receiving habits of a reader, he understands with little effort. When the symbols appear in less familiar sequences, comprehension is slower and less sure. And sufficiently improbably patterns seem like nonsense; they do not stand for anything in his experience (p. 419).

Taylor's Cloze Methodology

This writer is aware of Smith's work (1978) in psycholinguistics and of Rankin's efforts (1978, 1980) toward integrating cloze research findings into psycholinguistic theory. In this paper no attempts are made to circumvent or replicate this work. Rather, sources were reviewed which were in print prior to 1953, in order to capture the essence of Taylor's initial cloze methodology. In this respect Taylor's initial cloze methodology was clearly influenced by the work of Shannon, Weaver, Miller, and Osgood on several points:

1. Taylor clearly built the application of his procedure on dispositional mechanisms, noting it to be a task of comparing likenesses between the writer's intentions and the reader's anticipations.
2. In constructing his cloze passages Taylor included an adequate tolerance for redundancy, by deleting in the range of 10% - 20% of each of the passages (p. 423).
3. The testing methodology assumed a "corrections device" since students were not given time limits, and, thus could constantly forward correction data to refine the message.
4. Taylor removed words from a passage and in the place of each substituted blanks of a constant length, thus keeping the communications channel of a constant size.
5. Taylor may have further sought to maintain the size of the communications channel by using an every-nth word deletion system, but this is not entirely clear. Possibly, this only represented Taylor's method of sampling entropy/redundancy. He did specify that his deletion system was aimed only at measuring the readability of a given passage (1953, p. 420).

6. He instructed students to make use of entropy and redundancy by guessing the missing words from context.

There was, however, one area of concern which perplexed Taylor. In his series of pilot and experimental studies, Taylor included a pure random count for deletions on a passage along with an every-nth count. In comparing scores between the two, he found the pure random count to be the more efficient of the two systems (p. 423). Because the passage had only sixteen blanks, Taylor speculated that if the passage had more than sixteen blanks the results of the two deletion systems would have been similar (p. 425).

Taylor continued his work with the cloze. In 1956, he reported that cloze was a measure of learning. Here he had correlated pre-reading cloze and post-reading cloze scores to comprehension tests and to the subjects' Armed Forces Qualification Test scores. In 1957, Taylor reported that in general, the any-word deletion (every-nth) cloze yielded more reliable, more stable, and better discriminating results than did "easy" or "hard" cloze test forms. The latter cloze forms were constructed using a classification system in which the words from the test were classified as being hard or easy. The criteria for classification of each word was its function in context. In 1970, through a very lengthy discussion of the results of a pilot study, Taylor illustrated "latency" (time) and "focus" (subject interagreement) methods of cloze scoring. In this discussion Taylor was unable to establish the validity of cloze quantification as a logarithmic measure (p. 254). He did suggest, that the focus method (subject interagreement) had potential for constructing prose passages which would have vocabularly common to a given population (p. 257). From this it is inferred that passages constructed in such a manner would have validity for use in cloze research.

Other Research

In reviewing Taylor's research since 1953, it is apparent that he believed his original cloze testing methodology to be valid. At the same time it is apparent that he continued cloze research in order to further refine the method, as did others. For example, Bormuth (1965) found that accepting the exact answer seemed to yield the most valid score. Rankin (1978) echoes this point while Vaughan and Meredith question it (1978). Similarly, Thomas and Bridge (1980), in conducting cloze related research stated that the fifth-word deletion system was the most common system used, while Meredith and Vaughan challenged the efficacy of the automatic deletion system (1978).

There are numerous other studies which challenged the efficiacy of the traditional cloze testing methodology. At the same time it should be understood that the traditional methodology seems to have captured the essence of Shannon's and Weaver's communication system, Miller's speech and communications' theories, and Osgood's dispositional mechanisms. In doing so the traditional methodology fullfilled its promise as a valid measure of readability.

References

Bormuth, J. R. Validities of grammatical and semantic classifications of of cloze test scores. International Reading Association Conference Proceedings, 1965.

Hartley, R. V. L. Transmission of information. Bell System Technical Journal, 1928, 535.

Miller, G. A. Language and communication. New York: McGraw-Hill, 1951.

Meredith, K. & Vaughan, J. Stability of cloze scores across varying deletion patterns. In P. D. Pearson and J. Hansen (Eds.), Twenty-seventh yearbook of the National Reading Conference. Clemson, SC: The National Reading Conference, Inc., 1978, 181-184.

Osgood, C. E. The nature and measurement of meaning. Psychological Bulletin, 1952, 49, 197-237.

Rankin, E. F. Characteristics of the cloze procedure as a research tool in the study of language. In P. D Pearson & J. Hansen (Eds.), Twenty-seventh yearbook of the National Reading Conference. Clemson, SC: The National Reading Conference, Inc., 1978, 148-153.

Rankin, E. F. & Thomas, S. Contextual constraints and the construct validity of the cloze procedure. In M. L. Kamil and A. J. Moe (Eds.), Twenty-ninth yearbook of the National Reading Conference. Washington, D. C.: The National Reading Conference, Inc., 1980, 47-55.

Shannon, C. E. A mathematical theory of communication. Bell System Technical Journal, 1948, 27, 379-423.

Shannon, C. E. A mathematical theory of communication. Bell System Tech-

nical Journal, 1948, 27, 623-656.

Shannon, C. E. Prediction and entropy of printed English. Bell System Technical Journal, 1951, 30, 50-64.

Shannon, C. E. and Weaver, W. The mathematical theory of communication. Urbana: The University of Illinois Press, 1949.

Smith, F. Understanding reading: A psycholinguistic analysis of reading and learning to read (2nd ed.). New York: Holt Rinehart and Winston, 1978.

Taylor, W. L. Cloze procedure: A new tool for measuring readability. Journalism Quarterly, 1953, 30, 415-433.

Taylor, W. L. Cloze readability scores as indices of individual differences in comprehension and aptitude. Journal of Applied Psychology, 41, 1957, 19-26.

Taylor, W. L. and Waldman, I. N. Latency and focus methods of cloze quantification. In G. B. Schick and M. M. May (Eds.), Nineteenth yearbook of the National Reading Conference. Milwaukee: The National Reading Conference, Inc., 1970, 241-262.

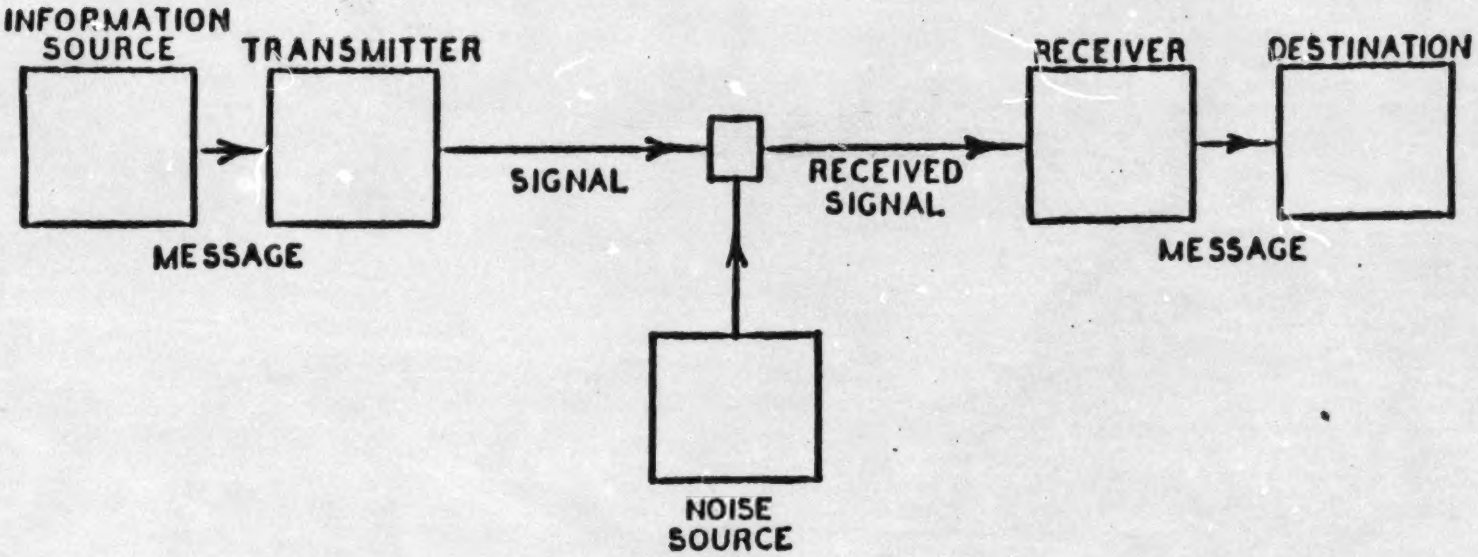
Taylor, W. Recent developments in the use of cloze procedure. Journalism Quarterly, 33, 1956, 42-48, 99.

Thomas, S. and Bridge, C. A comparison of subjects' cloze scores and their ability to employ macrostructure operations in the generation of summaries. In M. L. Kamil and A. J. Moe (Eds.), Twenty-ninth yearbook of the National Reading Conference. Washington, D. C.: The National Reading Conference, Inc., 1980, 69-77.

Vaughan, J. L. and Meredith, K. Reliability of the cloze procedure as assessments of various language elements. In P. D. Pearson and J. Hansen (Eds.), Twenty-seventh yearbook of the National Reading Conference. Clemson, SC: The National Reading Conference, Inc., 1978, 175-180.

Weaver, W. Recent contributions to the mathematical theory of communication. In C. E. Shannon and W. Weaver. The mathematical theory of communication. Urbana: The University of Illinois Press, 1949, 93-117.

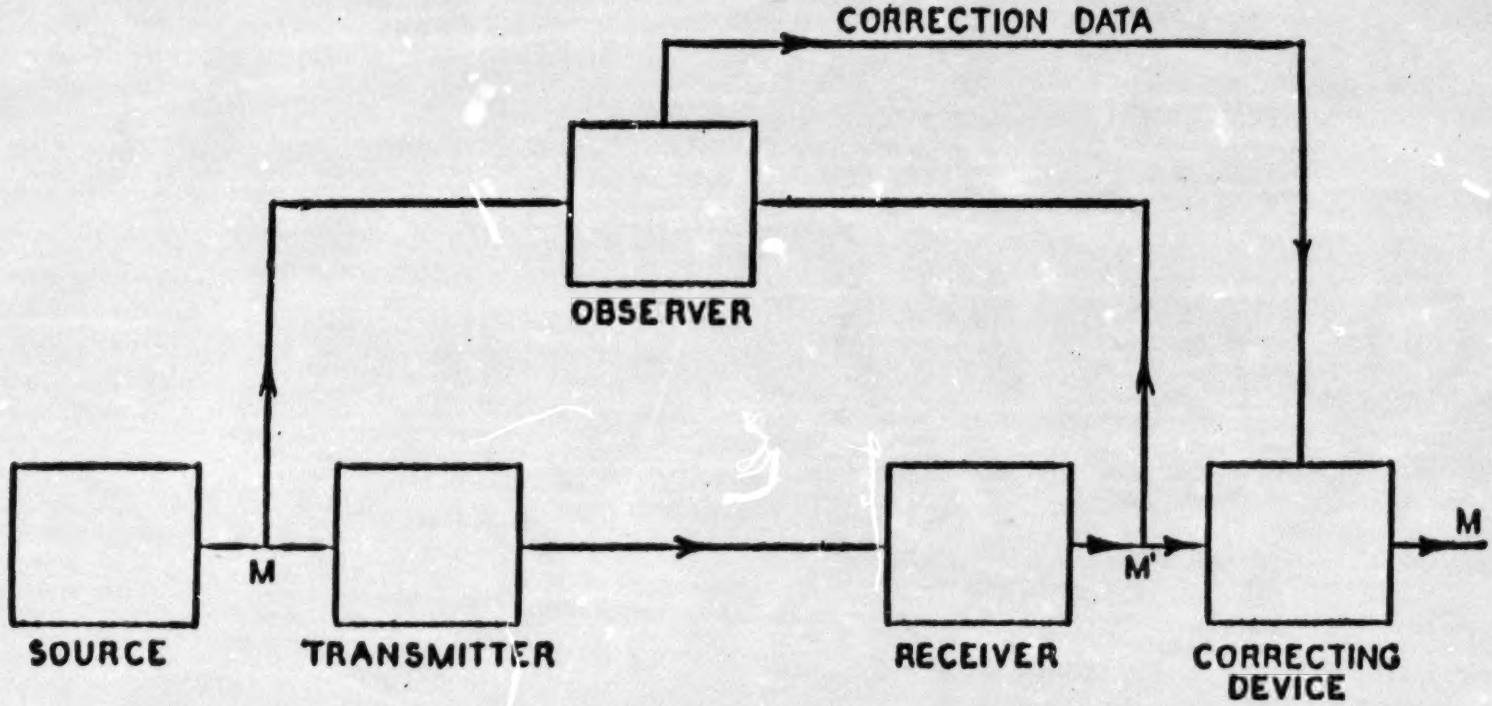
Figure 1



Schematic diagram of a general communication system.

From : Shannon, C.E. and Weaver, W., The Mathematical Theory of Communication, Urbana: The University of Illinois Press, 1949, 5.

Figure 2



Schematic diagram of a correction system.

From : Shannon, C.E. and Weaver, W., The Mathematical Theory of Communication, Urbana: The University of Illinois Press, 1949, 37.